

1. Type of data, samples, software, curriculum material and other material to be produced in the course of the project

Data that will be used in this study were collected during the *TARA Pacific* expedition and currently being analyzed by the different partner laboratories of the *TARA Pacific* Consortium. Collection of the data was largely funded by European sources (e.g., CNRS, Prince of Monaco, *TARA* Ocean Foundation, Genoscope and others mentioned in Gorsky et al. (2019)). As part of the consortium, our team was responsible for the collection and QC/QA of the inline system data (funded by NASA). New data products that will result from the proposed study are: pigment concentrations (from HPLC), optical proxies for chlorophyll, particulate organic carbon and size index, validate plankton images from FlowCAM, and diversity indices.

2. Standards to be used for data and metadata formats and content

TARA Pacific builds on the experience and success of *TARA Oceans* in discriminating data from over 35,000 physical, chemical and biological samples for public use, and providing the means to explore, access and assess raw and validated data sets (Pesant et al 2015, 2017). Controlled vocabularies describing sampling devices and sampling protocols, customized log sheets and unique sample identifiers for tractability were used throughout the expedition to capture provenance meta data (Gorsky et al., 2109). The structure of data sets and meta data will be consistent and interoperable with that of the *TARA Oceans* Expedition (Gorsky et al., 2019, described in detail by Pesant et al, 2015, 2017). If funded, we will contact BCO-DMO to register our project and discuss the procedures need to adapt metadata formats to BCO-DMO and link our project to the data repositories used by *TARA Pacific* (Figure 1).

3. policies for access and sharing, including provisions for appropriate protection and privacy, confidentiality, security, intellectual properties or other rights or requirements

Data generated as part of this project (optical inline measurements, HPLC pigments, validated FlowCAM images and diversity indices) will follow the NSF's 'OCE Sample and Data Policy' guideline and will be made publicly available within two years of the start of the project. Data generated by *TARA Pacific* partner labs will follow the *TARA Pacific* Data Usage Policy which follows that of *TARA Oceans* (Pesant et al., 2015). It is based on Open Science principle of early open access to raw and validated data sets, while offering those who contributed directly to *TARA Pacific* with exclusive rights on the use of data for a period of 12 months after data are published in public archives. During that period, anyone can use the data as long as the intended research does not compete with that engaged by the data originators. Anyone wishing to use *TARA Pacific* data while still under the 12-months moratorium period must contact the originators. All data will be published under the Creative Common Attribution 3.0 Unported and must be cited when used in scientific papers and presentations.

4. Policies and provisions for re-use, re-distribute and the production of derivatives

Once data will be published in public domains there will be no restrictions on the re-use, re-distribution and the production of derivatives.

5. Plans for archiving data, samples, and other research products, and for preservation of access to them.

Since the data used and generated by this project will be deposited in different repositories, we will work closely with BCO-DMO personnel to make the necessary links to all the repositories (Figure 1). We have already submitted the processed AC-S inline data to NASA's SeaBASS database. HPLC pigments data generated as part of this project will be submitted to BCO-DMO and SeaBASS. Currently there are no standards for data and metadata formats for data derived from imaging sensors such as FlowCAM and Zooscan. For the curation and annotation of images we use Ecotaxa, a web application developed by colleagues in France. Ecotaxa allows the curation of images with all the associated metadata and most FlowCAM images have been already deposited in ECOTAXA. PI Karp-Boss works routinely with this platform with imaging data collected for NASA funded projects (e.g., *TARA* Ocean Polar Circle, NAAMES and *EXPORTS*). She is part of an OCB working group aimed at establishing data standards and practices from taxon resolved phytoplankton images in effort to facilitate community-wide access to imaging data. The group works closely with representatives from BCO-DMO and SeaBASS regarding the desired path for the annotation and repository of digital plankton images. As the OCB working group establishes protocols for formats and repository of plankton images in national data centers, we will adapt our long-term plans for storage of images and data derived from this study. Derived metrics of community composition and biodiversity will be submitted to BCO-DMO.

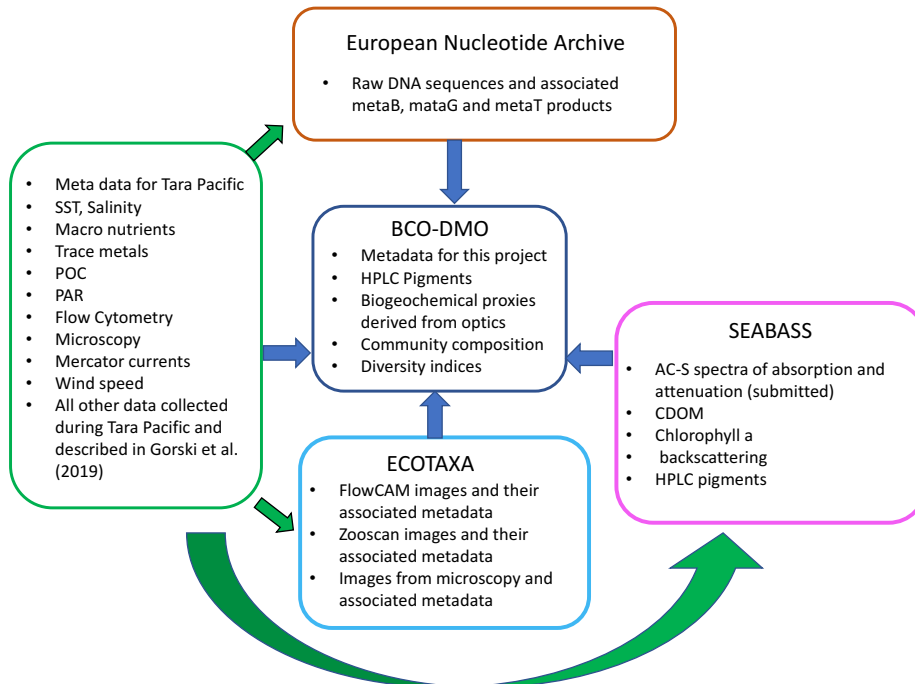


Figure 1. Archiving of TARA Pacific data